The Scholar's Forum Débuts

Richard C. Flagan
California Institute of Technology
flagan@cheme.caltech.edu

Anne M. Buck
California Institute of Technology
buck@library.caltech.edu

In recent years, scholars and academic librarians have begun to consider alternatives to the print-based model for research publishing. Fortuitously, two significant trends emerged almost simultaneously. While libraries faced canceling subscriptions in response to escalating journal prices and declining budgets, the creation of the World Wide Web made a ubiquitous electronic alternative to the print journal possible for the first time.

Among the first to act on the opportunities offered by the Internet and to set the model was Paul Ginsparg, a research physicist at the Los Alamos National Laboratory. In 1991, he developed www.arxiv.org/, a preprint server for the high-energy physics community. In the Ginsparg model, authors deposit preprints of their writings on a server, to which their colleagues around the world have open access. The initial appraisal that this would benefit only this very specialized discipline was soon refuted, as submissions from other areas within physics as well as related fields swelled the database to over 100,000 documents by 2000. Following on the success of the Los Alamos experiment, open-access preprint servers have been developed to serve a wide range of disciplines.

Ginsparg's preprint server augmented conventional journal publication by providing authors with a mechanism for distributing their writings long before they would appear in print; authors still submit their papers to get the stamp of approval that editors and peer referees provide. Publishers serving the physics community have accepted the preprint servers. Many of those serving other disciplines have resisted, fearing that open access to preprints will undermine their business interests.

Most other efforts to use the Internet for research publishing have extended the subscription journal model to the World Wide Web. Some journals have made their archival collections freely available through publisher servers, notably a number of offerings that are available through the Highwire Press. A few journals have made their entire holdings freely available through the Internet, again through publishers' servers. One organization, the Florida Entomological Society, has released its entire published record through a university server and makes the electronic version of its journal freely available immediately upon publication.

Debate on the impacts of open access to preprints or journal holdings on the future of scholarly publishing is heated. Most of the discussion focuses on the effects of open access on the business interests of society and commercial publishers. Some groups suggest that subscription journals can coexist with open access, but many argue that journals cannot survive this threat to their revenues.

The issues are real, and must be addressed, as editors consider the future of their respective journals. However, extending the present paradigm for dissemination of research reports is only one of many possible scenarios for the future of scholarly communication. In 1997, participants in a Conference on Scholarly Communication held at Caltech were asked how the research community would apply the capabilities of the Internet to research publications, if the print journal model did not exist. Dissemination suddenly becomes easy, but issues of quality, certification that a result meets the standards of the community as represented through an editor's acceptance of a manuscript after peer review, and preservation of the scholarly record must be addressed. The last issue is the hardest to address, since the only proven technology capable of preserving the literature is print on acid-free paper or, perhaps, clay tablets. No technology for storage of the electronic media has demonstrated archival capabilities. Unfortunately, many of the present print journals do not approach this severe standard. Only continued vigilance can ensure that the electronic media, or journals printed on low-grade paper, will remain accessible over the long term. Not only will technological solutions need to be developed, a number of legislated obstacles to preservation of the scholarly record will have to be addressed for this to happen.

The new media do present some exciting opportunities to change the way that scholars communicate their findings with their colleagues. Peer review is important, but it can be decoupled from the act of publishing. The scholar-editors who ensure the intellectual quality of the articles that appear in a journal, and the peer-referees whom these editors recruit to pass judgment on the works are volunteers who receive only modest support, if any, to cover the expenses of editing a journal. The largest journals may pay editors, but rarely at a rate that compensates realistically for the time and effort invested. Since the editors work independently, they need not be formally associated with the publishing operation.

The biggest beneficiary of the publication of a research manuscript is the author, whose reputation and standing in the field are enhanced, as colleagues

around the world are exposed to the work. The subscription model of scholarly publication places a gatekeeper between the author and the reader; only those who pay the subscription fee, either directly, through a library subscription, or through an access fee, can read the work. The Internet presents the opportunity to eliminate the gatekeeper by making access free to all who have interest in reading the work.

Internet publication is not free, however. If readers do not pay, who will? Since the author is the biggest beneficiary of scholarly publication, it seems reasonable that the author or the author's institution should assume responsibility for those costs.

The Conference on Scholarly Communication thus proposed a major change in how research results are communicated. Authors would bear the true costs of disseminating their work, but they would gain readers. The value of the end-state is clear, but how do we make the transition? A clear plan is needed to achieve the required technology developments as well as the change in author behavior.

The Scholar's Forum, http://library.caltech.edu/publications/scholarsforum.html, proposes a model for a collaborative approach to presenting scholarly findings, using networking technology to extend the traditional benefits of print journals. It addresses, but does not claim to provide a technological solution to, the significant issue of providing a stable archive of e-published works. Academic institutions would share responsibility for the support and maintenance of the platform and its contents; and the individuals and institutions that created the work would regain control over this important resource. At the same time, the Forum helps authors disseminate their findings in an environment free of subscription or licensing barriers.

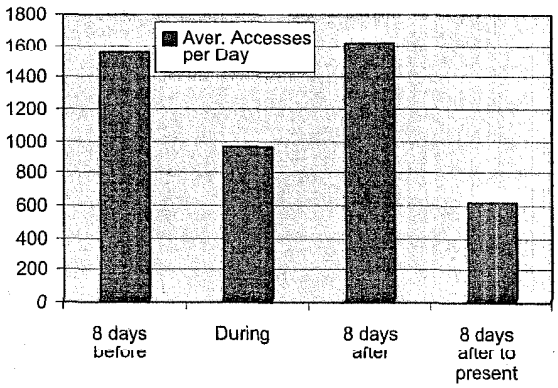
This model is, however, just a starting-point. The next step requires the key characteristic of much of the scientific research that the Forum ultimately seeks to publish, namely an experiment. Authors, particularly untenured faculty, are justifiably reluctant to commit their discoveries to an untested medium. Most potential editors lack the expertise to develop the networking tools that are essential to successful Internet publishing. Even if, as the Scholar's Forum proposal suggests, academic institutions commit to developing the necessary infrastructure, the combinations of the risks involved in those developments with the difficulty in attracting quality submissions are sufficient to deter even committed intellectual leaders. An experiment is needed that provides the combination of content and readers.

The Caltech Library System (CLS) and collaborators on the Caltech faculty have recently concluded a successful publishing experiment based on principles set forth by the Scholar's Forum. Demonstration of these principles could only be achieved by developing a substantial body of literature in collaboration with Editors whose primary interest was in the intellectual content of that literature, not in the development of Internet-based scholarly communication. On learning that the 4th International Symposium on Cavitation was to be hosted by Caltech in June 2001 (CAV 2001) (http://cav2001.library.caltech.edu), the CLS proposed to aid the conference organizers in developing the conference program by developing the necessary tools and assisting in the execution. Not only did all those involved gain valuable experience in e-publishing, but attendees and society members provided very positive feedback on the results.

Three important objectives set forth by the Forum were validated: (i) authors were responsible for the presentation of their work; (ii) document preparation costs remained with the authors; and (iii) the editor of the proceedings was able to receive and review abstracts of proposed papers electronically and respond quickly to authors. Accepted papers became instantly accessible for preview as soon as they were posted. Interestingly, access data revealed a substantial number of document downloads on the days prior to and following the conference; eight days after the conference the number of downloads leveled off (see Figures 1 and 2).

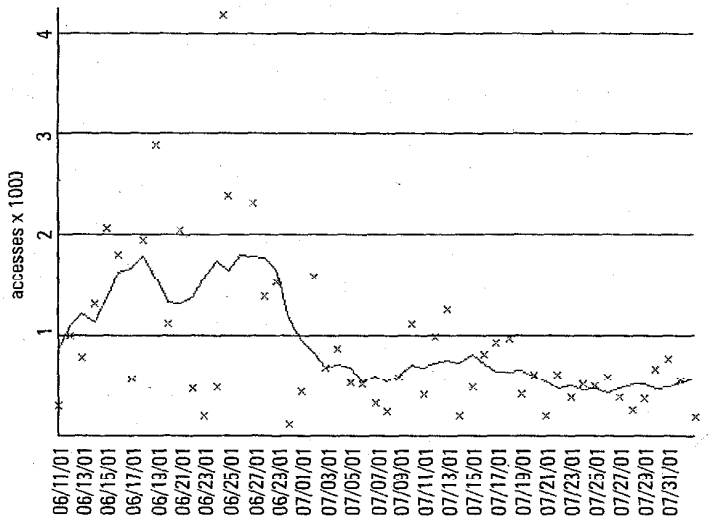
Other benefits became evident. The Cav2001 e-proceedings were available to participants and colleagues around the world; submissions came together in a standardized form regardless of country of origin. Conference attendees received the conference proceedings immediately, and avoided the added costs associated with printing, binding and mailing.

We also learned that e-publishing presents its own set of requirements, not the least of which is the need for a fast, persistent Internet connection coupled with a fundamental knowledge of computer systems. We chose PDF format over HTML to have displays and prints of high presentation quality. The technical infrastructure had to be able to adapt to international



Note: Roughly half of the accesses represent PDF downloads. Human users comprise 88% and search engine robots 12% of tota accesses based on browser type

Figure 1: Average accesses per day



Note: Each point on the line is the average access/day for a period of 7 days, comprising the current day +/- 3 days

Figure 2: Seven-day moving average

variables including language and differing tools for content creation and conversion.

From the outset librarians worked with the information technology (IT) specialists, and this collaboration proved highly productive. As the platform was being built, the IT people relied on the librarians to verify and validate important aspects of the way information would be managed. This partnership made it possible to reconcile differing points of view and ensure that the finished system would meet ongoing requirements for accessibility and usability.

Copyright is a central feature of the Scholar's Forum. Whereas most publishers take ownership of the intellectual property, the Forum seeks only an irrevocable license to make the work available over the Internet. The author retains ownership of the copyright, while providing open access to the work with proper attribution. Copyright was managed by requiring authors to register on the site, at which time they were given detailed information as to their rights. Acceptance of the terms of the copyright was made via click-through. The terms of use are also stated on the same HTML page as the download link to PDF, prior to downloading any document.

In the future, the society must address issues of creating a permanent e-archive and consider providing mirror sites or other distributed servers to ensure redundancy. It is difficult to envision how any single site can be considered archival. If one organization controls access to a work, that one organization can eliminate access too quickly for recovery. As stipulated in the Forum, the longevity of a document is directly proportional to the stability of its caretaker(s). Commitment to remain compliant with the Open Archives Initiative (OAI) will ensure that third parties may index content using an open protocol for metadata exchange. Proprietary formats of data storage must be avoided. In this way, papers will remain readily accessible to potential readers.

Four factors are contributing to the growth in networked scholarly communication: powerful searching capability based on OAI protocols; reduction in time from content creation to broad dissemination; potential for richer, more creative combinations of media formats (audio, video, etc.); and avoidance of commercial, cost-based barriers to access.

The CAV2001 proceedings project proves that the Scholar's Forum is a practical model, and that any conference proceedings can be published this way. This one experiment opens the door to a new model for scholarly communication that may be readily implemented and also meets the needs of authors, reviewers and attendees in a timely, effective manner. To become an accepted standard for disseminating research results, it is time to propagate the Scholar's Forum to other institutions including universities and professional societies; the essential prerequisite for the success of the Scholar's Forum model is the establishment of a collaborative consortium with sufficient resources, expertise and influence to generate a critical mass.